Project Summary/Abstract

The increasing accessibility and availability of genetic and genomic data has resulted in an influx of translational research to understand complex diseases. Multiple types of genetic and genomic data have been demonstrated to contribute to the underlying biological mechanisms of human disease. Genetic variants, gene expression, methylation and other genomic features often act in concert to effect change in cellular functions and downstream diseases. Genetic association methods traditionally have been used to analyze the relationship between variants and outcomes. This standard approach has been noted to have several limitations, including not elucidating the complex relationship between variants and diseases and not using potentially valuable information from other genomic data types.

This proposal aims at addressing these issues by developing integrative analysis methods for genetic and genomic data using (1) causal mediation analysis and (2) network analysis towards understanding the biological mechanisms underlying phenotypes of complex diseases. Specifically, causal mediation analysis provides an attractive framework for identifying the biological pathways that drive diseases. It proceeds by jointly analyzing multiple types of genetic and genomic data, where genetic variation is suggested to be mediated through genomic features. The method decomposes the overall effect of a variant on a disease outcome into the effect through the mediator, e.g. gene expression, and the effect through other biological mechanisms. Traditional mediation analysis for binary outcomes makes a rarity assumption that can be violated in complex diseases, such as asthma, which are common in certain populations. Estimators are constructed in this aim without imposing additional distributional assumptions for common binary outcomes.

Aim 1 proposes to use causal mediation analysis of genetic and genomic data in the setting of a common dichotomous outcome. In Aim 2, statistical inference for network analysis of genetic and genomic data is developed. Network analyses have also emerged as an integrative approach to characterize complex genomic associations. Features of a network of genetic and genomic variants can inform biological function. Current network methods treat edges as fixed and known, when in fact these relationships are estimated in the initial analyses. Given that they have uncertainty and error, it is important to estimate their error to ensure reliability and reproducibility of the results. Here, measures of error for network metrics are developed. The methods described in these aims will be applied to studies of asthma and chronic obstructive pulmonary disease. The success of this work will improve the ability to reproducibly detect relationships between various biomedical features using mediation and network analysis under many settings.